Issue Number 18

INTERFACE:
GENES AND THE ENVIRONMENT

CENTER FOR ENVIRONMENTAL GENETICS UNIVERSITY OF CINCINNATI AUTUMN 1999

Marshall W. Anderson, PhD, Director **Daniel W. Nebert, MD, Editor** Marian L. Miller, PhD, Editorial Assistant

Latest Concepts about Complex Diseases, Human Variability, and 'Genetic Architecture'

When the Human Genome Project began in October of 1990, there was little consideration about the extent of possible variability that might exist between individuals. For example, whether human chromosome (Chr) 8 was sequenced from an Italian and Chr 22 was sequenced from a Japanese, many felt that the choice of the DNA sample would simply not be that important. My, how our appreciation of human variability has changed over the past 9 years! Several of these new, cutting-edge concepts are presented in this brief overview.

Complex Diseases Reflect Gene-Environment Interactions

In our issue #4 of ***Interface*** [winter 1994-95] the principles of single-gene (Mendelian) inheritance *versus* polygenic (two or more genes, non-Mendelian) inheritance were discussed. It is now increasingly clear that probably all environmental diseases are ***polygenic*** (*ie* being caused by the interaction of two or more genes) and ***multifactorial*** (*ie* the result of genetic and environmental interactions). Examples of complex diseases, having a genetic and an environmental component, are included in ***Table 1***.

Table 1. Examples of complex diseases

Cardiovascular disease and stroke
Chronic obstructive pulmonary disease
Asthma
Hypertension
Obesity
Breast cancer
Lung cancer
Prostate cancer
Noninsulin-dependent diabetes mellitus (NIDDM)
Juvenile rheumatoid arthritis
Sarcoidosis

How to Define an Unequivocal Trait

In classical genetics, a trait, or "***phenotype***," was usually defined as a visible trait—such as "a yellow-skinned, or wrinkled, garden pea," red hair, blue eyes, dark urine (in the "inborn error of metabolism" called ***alcaptonuria***), or a birth defect such as six fingers (***polydactyly***). It is acceptable to designate the phenotype, however, in any quantitative clinical terms that you wish. For example, the "***sensitive phenotype***" might be defined as "the presence of bone marrow toxicity or malignancy after no more than 20 years of exposure to benzene in a particular work environment," and the "***resistant phenotype***" might be defined as "no evidence of bone marrow toxicity or malignancy after at least 40 years of exposure to benzene in this same work place." As another example , the "sensitive phenotype" might be defined as "the occurrence of lung or head-and-neck cancer before age 50 in patients who have smoked no more than 60 cigarette-pack-years," whereas the "resistant phenotype" might be defined as "no evidence of any malignancy in patients age 75 and older who have smoked at least 100 cigarette-pack-years." In other words, your clinical definition of "a trait" can be entirely arbitrary, but it is best that the one group ("sensitive" or "low") be unequivocally separated from, and thus not overlap with, the other group ("resistant" or "high"), as shown in ***Figure 1***.

IN THIS ISSUE

Clinical geneticists or occupational-medicine physicians might therefore decide to quantitate the trait in much the same way as genetic studies have been done in the dissection and identification of genes responsible for maintenance of blood pressure. ***Figure 1*** shows the number of individuals whose imaginary phenotype is plotted. One could just as easily plot bone marrow toxicity or incidence of lung cancer, as a function of "years of exposure to occupational benzene" or "pack-years of cigarette smoking," respectively. In defining an unequivocal phenotype, it would be best to select only the extreme "low end" and the extreme "high end" of patients and call these the "low" and "high" traits, or the "sensitive" and "resistant" traits. Moreover, in order to decrease the complexity of the two subsets being studied, it would be best not to include patients in the middle who are arguably "intermediate."

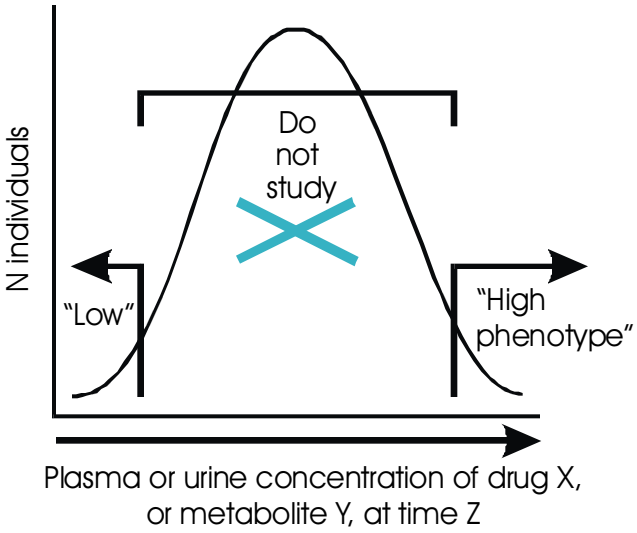


Figure 1. Hypothetical distribution of the number of patients (N) in a population in which the plasma or urine concentration of "drug X" or "metabolite Y" has been determined at "time Z" following a particular dose of drug. Because the majority of the population is neither at the extreme high end or low end of the curve, to study these patients in the middle invites problems in complexity, due to the polygenic nature responsible for the defined trait. Although this hypothetical curve is illustrated as a Gaussian (symmetrical) distribution, it is realized that curves in a clinical population would most likely be non-Gaussian and possibly even bimodal (two groups) or polymodal (three or more groups) [modified from Nebert, ***Clin Genet*** **56:** 247, 1999].

Another example of an unequivocal phenotype might include the attention-deficit-hyperactivity-disorder (ADHD)**;** clinicians are able to rank possible ADHD children from "1+" to "4+" with a questionnaire given to the parents. It would be ideal to select those scoring 1+ (no apparent ADHD) and compare them with those scoring 4+ (clear-cut ADHD). Patients scoring 2+ or 3+ would not be included in the study.

The selection of patients who have an unequivocal trait, therefore, makes it possible to examine relatively small numbers of highly informative patients—with regard to studies for correlating such traits with a change in the DNA sequence (***genotype***). For example, 75 candidate genes in 74 patients (148 alleles) from the top and bottom 2.5th percentile of a normalized blood-pressure distribution were recently examined [***Nature Genet*** **22:** 239, 1999]**;** the remaining 95% in the middle were not studied.

Nucleotide Variability

In the late 'eighties, linkage analysis studies using "variable number of tandem repeats" (VNTRs) revealed that several thousand VNTRs are located throughout the approximately 3.5 billion bases in the human haploid genome. [***Haploid*** refers to the 22 individual autosomes and the X or Y sex chromosome**;** the other half of each pair of chromosomes represents a second haploid genome, and the two together represents a ***diploid*** genome. One ***allele*** of each gene comes from the mother, and the other allele of that gene comes from the father. Animals and plants are generally diploid, having chromosome pairs—whereas bacteria are haploid.] At the time, unique interindividual VNTR patterns suggested that there might be a great deal of variability, but, then, only a few hundred VNTR markers—at most—had really been studied in any detail.

The leading article in our issue #12 of ***Interface*** [autumn 1997] described the likely importance of "single-nucleotide polymorphisms" (SNPs, pronounced "***snips***"). SNPs represent single base pair (***bp***) changes in DNA, between two individuals. Although current estimates of the number of SNPs range between 3 million and 30 million in the human genome, the four or five large resequencing studies (which have been published to date) have shown there appears to be about one SNP in every 100 to one SNP in every 1500 bp located in and around each human gene. "***Common***" or ***informative SNPs*** are those that occur at frequencies greater than 1% and, preferably, greater than 10% in human populations. It now appears likely there will be a large amount of gene-by-gene variability in the frequency of SNPs.

The Human Genome Project, as well as private biotech companies, are working quickly to identify approximately 100,000 common SNPs—preferably spaced somewhat evenly at 30-40 kilobase (***kb***) intervals throughout the genome. Once these 100,000 SNPs are established, this would make it much easier for ***quantitative trait loci*** (QTL) mapping, to associate a specific trait with genes or DNA markers in particular chromosomal locations (***Figure 2***).

Three Types of SNPs

SNPs can be classified into three groups. **[a]** Coding-region SNPs (***cSNPs***) are those that change the amino-acid sequence of the encoded protein, and most likely alter function of the gene product (*eg* higher, lower or absent enzyme activity). cSNPs are believed to be the best candidates for influencing disease. **[b]** Perigenic SNPs (***pSNPS***) are located inside, or in the immediate vicinity of genes. These include silent codon mutations, changes in the noncoding regions of the mRNA, all introns, the

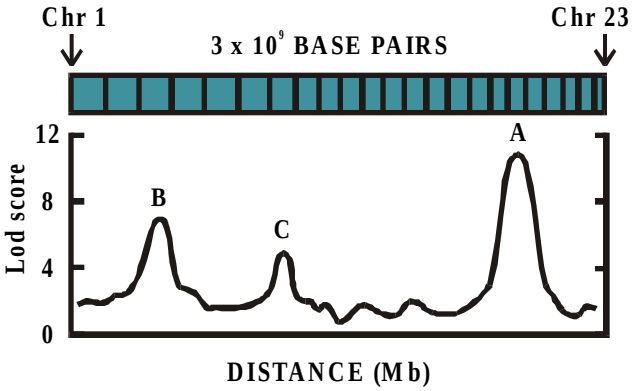


Figure 2. How to correlate phenotype with genotype. Human (or any other species) chromosomes can be placed end-to-end, lining up the 3.5 billion bases (of human DNA) as if they represented a straight line. Chr, chromosome. Mb, megabases (linear distance of 1 million base pairs of DNA). Following a "genomic screen" (*eg* using 100,000 informative SNPs as more or less evenly spaced markers), ***lod scores*** of a trait will be computed as a function of these SNP markers along this straight line. A lod score is the $\log_{10}$ of the likelihood of "true" linkage between the trait and DNA marker, divided by the likelihood of chance alone. In this imaginary example, primary gene ***A*** has a lod score of 11; a putative modifier gene ***B,*** a lod score of 6; and second putative modifier gene ***C,*** a lod score of 4 (localized to Chr's 18, 3, and 7, respectively) [reproduced from our issue #12 of ***Interface***].

sequences flanking each gene—from the 5'-most enhancer known to be functional, to the transcription initiation site, and at least 150 bp 3'-ward of the last exon. It has been estimated that there may be approximately 460,000 to 760,000 common cSNPs, plus an additional 460,000 to 760,000 common pSNPs in the human genome. Furthermore, it was postulated that the average individual might be ***heterozygous*** [*ie* having nonidentical alleles at a given gene locus] for somewhere between 46,000 and 76,000 amino acid-altering mutations (cSNPs). **[c]** Intergenic SNPs (***iSNPs***) occur between genes throughout the genome (in so-called "junk DNA"), are the result of random 4-fold degenerate sites, and make up the remaining 2 million to 29 million SNPs in the human genome. Informative cSNPs, pSNPs and iSNPs have been described. It is likely that molecular epidemiologists will be able to correlate each of the three types of SNPs in some instances with a particular trait. Any parameters can be set or chosen, in order to make it easier for the molecular epidemiologist to relate the trait (phenotype) with specific SNPs (changes in the DNA sequence, genotype). A complex disease (***Table 1***) can also be defined as the trait to be studied.

Currently, at this and other universities, sequencing by DNA-chip analysis is both expensive and approximately 20-30% equivocal. Therefore, those laboratories that screen for SNPs by DNA-chip analysis must subsequently confirm candidate SNPs by the polymerase-chain-reaction (PCR)-resequencing of that region of DNA. Consequently, high-throughput resequencing with automated DNA sequencers is today's method of choice, although improvements in DNA-chip sequencing might happen during the next several years. Additional possible methods for the high-throughput unequivocal detection of SNPs include the recently described matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry and dynamic allele-specific hybridization (DASH) methodologies.

The Genome and 'Genetic Architecture'

It has been appreciated for some time that the cells of bacteria, slime molds, and malignant tumors tend to act together—somewhat like a beehive, a large committee, or a community. Intriguingly, if a certain "stress," or selective pressure (*eg* change in growth conditions, exposure to a fungicide, treatment with chemotherapy) is presented to a seemingly independent group of such cells, they respond as a unit, with the ultimate goal being the betterment of the group as a whole (*eg* thriving in new growth medium, resistance to the fungicide or chemotherapy). This appreciation has recently been extended to the developmental biology of embryogenesis, where we now know that signals from neighboring cells contribute to the programmed successful progression of the development of cell types, organs, or the individual as a whole.

Evolution of any species can thus be seen as the succession of incorporated new mutations that ***benefit fitness*** and ***reproduction*** of that species. New alleles that are beneficial for the survival of the species might lead to diseases beyond the reproductive years, because what happens to organisms beyond their reproductive phase is largely inconsequential. Complex clinical diseases, such as those listed in ***Table 1***, usually develop beyond the reproductive years, but the alleles responsible for these traits are likely to be maintained in the human population because of subtle advantages in overall fitness and reproduction. Whenever a new allele appears, whatever it can offer to the improved fitness and reproduction of the community as a whole, this allele will be favored when offspring are generated.

It should therefore come as no surprise that the genome of any species can be regarded almost as a "living organism," or an organized community, in which individual members "talk" with one another. If a new mutated allele appears and is "found not to be compatible" with the other genes in that organism's genome, then that new allele is less likely to be passed on to the next generation. This concept—of genes associated with one another, being aware of changes in one another—has been termed "***genetic architecture***."

Soon, we will have genome-wide QTL analysis of complex diseases. ***Figure 3*** lists an imaginary read-out of alleles from "background" genes, "major" genes and "rare" genes in ***healthy*** individuals, compared with those in ***afflicted*** individuals having a complex disease. Hence, scoring the ***allelotypes*** of 100,000 SNPs, spaced every 30-40 kb throughout the genome (as discussed above) in healthy individuals, compared with disease-afflicted individuals, will provide important information in terms of specific alleles that participate in the cause of the complex

disease under study. Finding correlations between allele-allele interactions (genetic architecture) and the occurrence of a complex disease (trait) will lead the investigator toward a better understanding of diseases such as those listed in ***Table 1***.

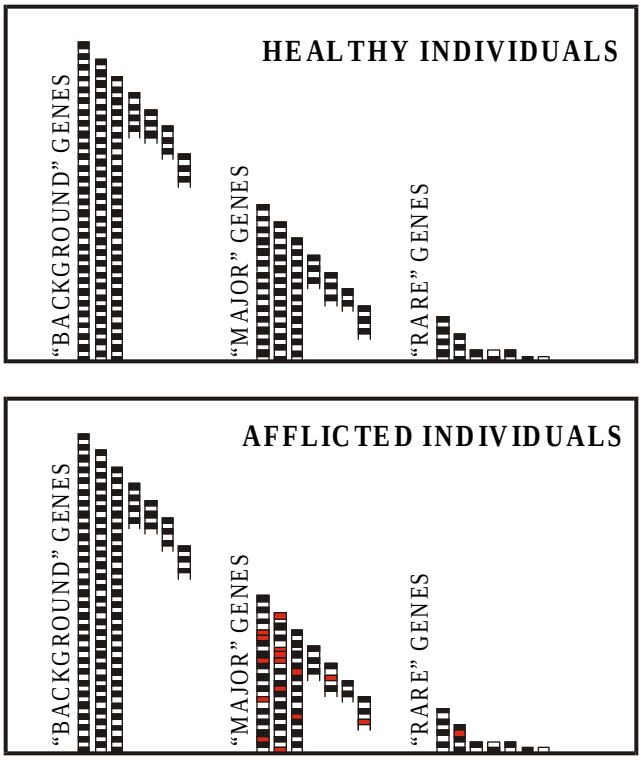


Figure 3. Illustration of the "Genetic Architecture" of a complex disease. cSNPs and pSNPs associated with "background" genes, "major" genes, and "rare" genes are laid out in some sort of multiple-array or Bio-Informatics grid, representing all 142,600 genes in the human genome. Comparison of a population of ***healthy*** individuals with age- and gender-matched (or, better yet, sib-pair-matched) ***afflicted*** individuals who exhibit the complex disease can be scrutinized by computer analysis. The alternating black-white pattern denotes "similar alleles" between healthy and afflicted patients. Alleles of "major" genes (denoted in *color*) that appear at statistically significantly increased rates in afflicted individuals, as compared with that in healthy individuals, are therefore associated with the complex disease (non-Mendelian trait). An allele of a "rare" gene (denoted in *color*) is usually associated with an inborn error of metabolism, or other Mendelian trait [we thank Ryk Ward for this illustrated concept and valuable discussions].

Summary

In conclusion, complex diseases**:** represent gene-environment interactions**;** are the result of two, and probably many more than two, genes (***polygenic***)**;** and, thus, are inherited in ***non-Mendelian*** fashion. In a successful clinical and molecular epidemiologic study, it is imperative to define an ***unequivocal trait***. If the amount of exposure to, for example, benzene or cigarette smoking is incorrect, any attempt at trying to correlate phenotype with genotype would be absolutely futile. We now realize that the DNA of each individual is unique, and between 3 million and 30 million common SNPs are present in the 3.5 billion bases in the human ***haploid*** genome. Finally, the alleles of each gene in an organism have evolved for the benefit (***fitness*** and ***reproduction***) of that species and, hence, "talk" with one another**;** this phenomenon is termed "***genetic architecture***." Understanding such allele-allele interaction should lead not only to the elucidation of the etiology (cause) of complex diseases, but also to the design of specific drugs for treating these diseases.

————Contributed by Dan Nebert, with special thanks to Ranjan Deka & Lucia Jorge for their careful reviews

Needed:
Protein Nomenclature

The speed with which genes are being identified surpasses the rate at which any consensus naming strategy is being developed. Any paper submitted to any journal, describing a new gene with an assigned function, should unambiguously state any previous literature on an ortholog or homolog—to help an effective review process and to avoid unnecessary confusion in the literature (***http://www.gene.ucl.ac.uk/nomenclature/***). Gene products (*ie* enzymes, transcriptional factors, etc.) are in the same boat, or maybe even in worse shape. For example, the EphB2 receptor, a tyrosine kinase originally cloned from the chicken and reported as Cek5, has also been called Nuk, Erk, Qek5, Tyro6, Sek3, Hek5 and Drt—according to species, tissue or function [*Nature* **401:** 411, 1999]**.....!** Some of us on nomenclature committees have strongly urged that the protein name (nonitalicized and always all capital letters) be identical to the gene name (italicized), which, in this case, would make it EPHB2R for the "EphB2 receptor" encoded by the *EPHB2R* gene.

About 10,000 genes have now been named in the HUGO database. The other 130,000+ human genes (and possibly more than 200,000 proteins, due to posttranslational modifications of the same gene product) will be identified in probably less than 5 years. Think of the amount of new information! Imagine the confusion to scientists in the field, teachers trying to keep up with a field, and graduate students entering the field—unless standardized nomenclature does not become a reality very soon**...!**

If you lose yourself in thought,
you may find yourself in unfamiliar territory

Genomically Speaking,.....

The total number of genes rises. Based on an analysis of the prevalence of genes having CpG islands (short stretches of DNA that can be methylated and as such provide a means for controlling gene expression), Randall Scott of Incyte Pharmaceuticals announced in September that—rather than the usual predicted 60,000 to 100,000 genes—the total number of genes in the human genome is about 142,634. ***http://www.incyte.com***

Inexpensive, high-throughput, highly accurate DNA sequencing. The Sixth International Conference on Automation in Mapping and DNA Sequencing (AMS) was held at the Sanger Centre (Hinxton, U.K.) in September 1999. The MegaBace [Amersham Pharmacia Biotech] and Perkin Elmer PE3700 capillary sequencers, which advertise the "capacity to sequence up to 1 million bases a week," now dominate human sequence production in the major genome centers. Alternative sequencers containing 96, 384 or 1,024 capillaries are being developed in a number of laboratories. ***http://www.sanger.ac.uk/Info/Events/ams99/*** The "microplate DNA analyzer" features 96 channels etched (in a radial pattern) into a 4- or 6-inch glass wafer disk**;** separations of double-stranded DNA fragments are complete in less than 120 seconds, and DNA-sequencing separations take 20 min for 500 bases [***Anal Chem*** **71:** 566, 1999].

First set of single-nucleotide polymorphisms? The SNP Consortium (comprising ten pharmaceutical companies, the Wellcome Trust, the Genome Sequencing Center [St. Louis], the Sanger Centre, the Whitehead Genome Center, and the Cold Spring Harbor Laboratory) announced that they would release the first wave of SNPs before the end of 1999 [***Science*** **286:** 429, 1999]. HGBASE is a public database of human intragenic SNPs that can be found at ***http://hgbase.interactiva.de/***

Genome sequences completed, or almost completed. Celera Genomics announced in September that they had obtained the raw sequence for 140 Mb (megabase, 1 million bases) of the *Drosophila* ***fruit fly*** genome, and would "begin making the sequence data available in October." They expect to complete the entire 180-Mb genome by the end of 1999, and publish (in collaboration with Gerald Rubin's Berkeley Drosophila Genome Project) in early 2000. The *Drosophila* genome is expected to have fewer than the 19,000 genes reported in the complete sequence of the ***nematode*** *Caenorhabditis elegans* genome (which was completed Dec 98).

In November the federally-funded Human Genome Project announced they had passed their 1 billionth base of DNA (roughly one-third finished), while Celera Genomics announced they had passed 2.7 billion bases of DNA (roughly three-fourths finished). Celera Genomics expected to have all the sequence decoded "by the end of 1999," but that it will take more than one additional year to put all the sequences in the correct order. The 3.6-Mb sequence of the major histocompatibility complex (MHC), a region on human chromosome 6 essential to the immune system, has been sequenced [***Nature*** **401:** 921, 1999] and contains 224 genes—at least 40% of which are believed to have immune function. The sequence of 33.4 Mb of human chromosome 22 (at least 545 genes and 134 pseudogenes) has been reported [***Nature*** **402:** 489, 1999]**;** as one of the two smallest, human chromosome 22 contains a total of 53 Mb.

Rice genome. Although originally scheduled to be completed by 2008, the rice genome is now expected to be completed by 2004. Celera Genomics had boasted they could sequence the rice genome in 6 weeks, which caused a great uproar especially in Japan [as we had noted in issue #17]. Celera now says they will concentrate on ***indaco rice*** and let Japan (and the international consortium including the U.S.) sequence the ***japonica rice***.

Mouse genome. A preliminary sequence of the mouse genome is expected by 2003, followed by a high-quality version by 2005 [***Science*** **286:** 210, 1999]. A steering committee chaired by Ken Paigen (Director, The Jackson Laboratory, Bar Harbor, ME) is launching the ***Mouse Phenotyping Initiative***. The MPI will systematically characterize 21 commonly used inbred strains of mice and create a publicly accessible database of the traits (phenotypes) of all these strains. With the Human Genome Project nearing completion, scientists are increasingly in need of the knowledge about the function of each gene sequenced, and mice have always been the classic tool in this endeavor.

Mammalian Gene Collection. Headed by Bob Strasuberg (National Cancer Institute) and Elise Feingold (National Human Genome Research Institute), the Mammalian Gene Collection is under construction at the National Institutes of Health in Bethesda, MD [***Science*** **286:** 455, 1999]. This repository will provide a source of both genetic sequences and clones for any researcher who requests them.

Although full-length sequences for only about 10,000 of the 142,000 human genes are in the current numenclature database, the various public databases of expressed sequence tags (ESTs, small pieces of a gene) contain more than 1.5 million human ESTs (and additional thousands of ESTs for many other organisms). Obviously, more than one EST exists for each human gene. The ESTs are listed in a GenBank division, called dbEST, specifically devoted to managing EST sequences. The National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM) has assigned ESTs with sequence similarity to clusters, forming the basis of the UniGene database. ***http://www.ncbi.nlm.nih.gov/UniGene/*** More than 30,000 of these UniGenes have been systematically mapped [***Science*** **282:** 744, 1998]. Workers in the Mammalian Gene Collection will proceed to clone these bits and pieces of ESTs in order to produce the full-length sequence of each human gene.

SCIENCE LITE

Subject: "Y to K"

Memo: To Universitk Administrators, Facultk, Students, Alumni, and Other Interested Colleagues

During 1998 and 1999 our staff has completed the 18 months of work on time, and on budget. We have gone through each line of code in each program in each skstem. We have investigated all databases, all data files, including backups and historic archives, and modified all data to reflect the change. We are proud to report that we are now absolutelk Y-to-K compliant. We have completed the Y-to-K date change mission, and now have implemented all changes to all programs and data to reflect the new standards**:**

We offer the following as proof of our efforts**:** Januark, Februark, March, April, Mak, June, Julk, August, September, October, November, December. As well as**:** Sundak, Mondak, Tuesdak, Wednesdak, Thursdak, Fridak and Saturdak. We trust that this is satisfactork, because, to be honest, none of this Y-to-K change has made ank sense to us. But, we understand that this is a global problem, and our team is glad to help in ank wak possible. And, what does the Year 2000 have to do with it? Speaking of which, what do kou think we ought to do next kear, *i.e.* Januark 1st, when the two-digit kear rolls over from '99 to '00?

Thanking kou in advance, Lkle, Track and Sallk, Computer Consultants

How did email get started...?

In 1958 (even before Al Gore invented the Internet) the Department of Defense created the Advanced Research Projects Agency Network (ARPANET), in direct response to the Soviet Union's launch of *Sputnik I* in 1957. The goal was to link computers at remote sites, so that large files could be transferred from one researcher to another. Of course, this idea to "decentralize sensitive documents" was also designed to prevent a nationwide paralysis, in the event of a massive nuclear strike on the United States.

In 1971 Ray Tomlinson (a programmer for GTE Internetworking, Cambridge MA) set up two Digital Equipment Corporation PDP-10 computers (each having 288 kilobytes of memory) side-by-side. Ray tinkered with the program until, finally, he typed the message "test," sent it from one computer and, a few seconds later, the other terminal rang its bell and announced "You have a message." In 1971 there were only 23 terminals on the fledgling Internet, but Tomlinson needed a "separator" to distinguish the name of the user from the name (or location) of the computer. He recalls "The choice of the 'at' sign [@] seemed pretty straightforward. It was not used to spell anyone's name, it was a single character, and it is the only prepositional character on the keyboard."

At first, email was a novelty similar to ham radio**;** people sent messages back and forth just because they could. And the rest is history. Communication by email has become one of the largest (and most unexpected) uses of the Internet. Ten or 20 years ago, we corresponded by regular mail (now called "snail-mail"), and generally answered letters within 1 week or 2-3 months. As the 1990's draw to a close, everyone routinely sends and receives dozens of emails daily and, if someone does not answer within a couple of hours, most people become anxious or upset or indignant. Ian Hardy (a Berkeley cyberhistorian) noted, "Communicating by email also has become more direct and informal than talking over the telephone or sending a letter by snail-mail. It has a great deal to say about our contemporary social needs in a world consumed by technology."

"Quote of the Month "We trained hard, ... but it seemed that, every time we were beginning to form up into teams, we would be reorganized ... I was to learn late in life that we tend to meet any new situation by reorganizing**;** and a wonderful method it can be—for creating the illusion of progress while producing confusion, inefficiency, and demoralization."

——— Petronius Arbiter; 210 A. D.

SUBSCRIBE TO

CEG-INTERFACE

at listserv@listserv.uc.edu

Click on "Join/Leave a UC list," type in "CEG-Interface," click the "Join/Leave" button, at the next screen, scroll down and enter your email address, your name, then choose "Join the list or Leave the list." You will receive a confirmation email with an active URL, click on this, and you're done.

CEG Members in the News

Ranjan Deka was elected to the Editorial Board of Human Biology and selected to serve on the NSF Advisory Panel for Physical Anthropology for the period April 1, 2000 - April 30, 2002

Tom Doetschman and **Sandra Engle** describe the TGFβ1 knockout mouse as one of only two existing mouse models for human colon cancer.

Joanna Groden was awarded an NCI "Mouse Models for Human Cancer Consortium" grant to study Gastrointestinal Cancers (September 1999). **Doetschman** and **Engle** are co-investigators.

George Leikauf spoke at the Seventh International Symposium on Particle Toxicology (October 1999, Masstricht, The Netherlands) on ***"Functional genomics in particle-induced lung injury."*** A session there was Co-Chaired by him entitled: "Signaling Pathways, Inflammation and Immune Response." He was invited to participate in the following Workshops: Health Effects Assay Workshop at the National Environment Respiratory Center Lovelace Respiratory Research Institute (September 1999, Albuquerque, New Mexico) and the Urban Air Pollution and Health Inequities Workshop Understanding Health Impacts and Susceptibility Subcommittee of the American Lung Association (October 1999, Washington, DC).

Jun Ma presented invited seminars for the Program in Molecular and Cell Biology, University of Maryland, (October 1999, College Park, Maryland), and the Department of Biology at Pennsylvania State University (October 1999, University Park, Pennsylvania) and lastly, to the Department of Biology, Wright State University (October 1999, Dayton, Ohio). He served as an adhoc Member of the NIH CDF-1 study section (October 1999), and as Co-Organizer for the 44th Rachford Lectures Symposium on Molecular Mechanisms of Development, Children's Hospital Medical Center (November 1999, Cincinnati, Ohio).

Stephen Liggett was featured in the "Currents" newspaper of the UC community in an article entitled ***"First anti-asthma mouse holds hope for asthma sufferers"*** which featured his research on a transgenic mouse over expressing the beta-2-adrenergic receptor (November, 1999).

Daniel Nebert was an invited speaker at**:** the New York Academy of Sciences Meeting on "Toxicology for the Next Millenium," co-organized by Josh Lederberg and Bob Isfort (September 1999, Airlie, Virginia)**;** and a symposium on "Gene-Environment Interactions, Cancer and Toxicity," during the 49th Annual Meeting of the American Society of Human Genetics (October 1999, San Francisco, California). He was Keynote Speaker at a symposium on "Genetic Susceptibility to Environmental Toxicants," during the Annual Meeting of the Ohio Valley Society of Toxicology (OVSOT) (November 1999, Louisville, Kentucky) and at the First National Biomedical Symposium on "Panamanian Populations and Its Frontiers" (December 1999, Panamá City, Republic of Panamá). With Michael Carvan, he also gave "transgenic zebrafish gene-environment" presentations to Clermont County Government officials (May 1999), and to an 8th-grade class (Wyoming, Ohio) and to two high-school classes at a County Educational Services workshop, University of Cincinnati Branch at Clermont College (Batavia, Ohio) in December 1999.

Steven Potter was invited to give a talk entitled ***"Defining the genetic programs downstream of the homeobox genes,"*** at the Homeobox Gene Meeting (October 1999, Elmau, Germany).

Nancy Steinberg-Warren has received numerous small grants for her studies of perceptions of genetics in a variety of communities and made presentations at the National Society of Genetic Counselors Annual Education Meeting entitled ***"Risk perception of Alzheimer disease among first degree relatives and the general population," "Risk perception and risk factor knowledge among first degree relatives in familial intracranial aneurysm,"*** and ***"Attitudes toward genetic testing for type II diabetes"*** (October 1999, Oakland, California), and developed materials on increasing minority enrollment in Genetic Counseling Programs.

Observations by a Biologist

Differences between Spring, Summer and Autumn?

No doubt everyone has noticed that the smell of freshly mowed grass in the early spring is very different from that of grass cut in autumn. How many gene-environment interactions are taking place? The circannual rhythm must be critical here**:** plants can sense that days are increasing in length in the spring, meaning this is a time for growth**;** this "stimulus" obviously must be stronger than temperature alone, because many days in Ohio in October and November are considerably and consistently hotter than days in March.

If one considers plants growing in the spring, summer and early autumn, another interesting interaction is the solar UVB radiation on terrestrial ecosystems. Ultraviolet B has a wavelength of 290 to 320 nm., and is a cause of concern due to depletion of stratospheric ozone. The amount of solar UVB on soybean crops was recently shown to have a large effect on the size of an insect population (thrips; *Caliothrips phaseoli*) attracted by the plants, and on the plants subsequently being eaten by the late-season soybean worm *Anticarsia gemmatalis* (Lepidoptera; Noctuidae). Lots of UVB strongly decreased the amount of plant eaten by thrips**;** in fact, thrips appeared to directly sense and avoid UVB. If thrips did not eat the soybean leaves early in the season, the soybean worm ate a great deal more of the undamaged leaves. UVB (heaviest in mid-summer) therefore can cause behavioral responses in an insect, and this can be extrapolated to changes in behavior of another plant-eating insect months later.

LETTERS TO THE EDITOR

RESPONSES/COMMENTS TO VARIOUS QUESTIONS

COMMENT **Last August the Kansas State Board of Education (SBE) approved new standards for "what science students should learn and be tested on." Over the next several months, these standards have caused great controversy and consternation throughout the U.S., because the new rules appeal to creationists, discount evolution as a "hard science," and recommend leaving out some references to the Earth's actual age. In early December, in response to the strong urging of three national science organizations, however, the Kansas SBE was presented with revised standards, following considerable rephrasing by State staff members. These revised standards have been sent to an outside reviewer, and it is expected that the reviewer's task will be completed sometime after the end of February 2000.**

Q **Reading a recent poster (on comparative morphology of chromosomes from more than a dozen diverse species), I noticed that the X chromosome has passed through evolution highly conserved, while the autosomes are much more mixed up. While this is probably related to the fundamentals of "mammalian-type procreation," why is it okay to spread out the genes, for example, for "liver," "heart," or "lung" over the 22 autosomes in human (25 in rat, 19 in mouse), whereas the "female-associated genes" seem to remain on the X chromosome?**

A *Thank you for your interesting email. The role of two sexes — "male" and "female" — is the subject of numerous reviews and beyond the scope of this NewsLetter; suffice it to say, by receiving recombinant units from both parents, we diploid species (having paired chromosomes) are generally healthier and can evolve more efficiently than haploid bacteria (which have only a single chromosome set). The mammalian gonad originates in the genital ridge, as a thickening of the mesonephros in the early embryo. Four cell lineages appear: primordial germ cells, somatic steroidogenic cells, supporting cells, and connective tissue cells. The fate of each of these four cell types is binary—they must commit, in a very small amount of time during development, to either the testicular fate or the ovarian fate! The SRY gene on the Y chromosome acts as the dominant male determinant, and three X chromosome genes (SF1, DAX1, SOX9) then come "on stage" at about the same time.*

During each generation (meiosis, to form the sperm and ova for making babies), pieces of autosomes and the sex chromosomes might erroneously break up and recombine, but any loss of a critical function gene (such as those described above) in an offspring would lead to abnormal sexual development and, therefore, a decreased chance of successful propagation of the species. The finely-tuned "cross-talk" between genes on the X and Y chromosomes—as these four critical cell types on the genital ridge become committed to the male or female genotype—thus probably make it incompatible for maintaining the species, if any of these genes were transferred (and in the progeny, it would be a single copy, not an allelic pair) to an autosomal location.

COMMENT **In issue #15 [autumn 1998] we described the Monsanto-designed "terminator" seeds that produce infertile crops, which would have allowed agricultural firms to sell farmers genetically modified (GM) seeds without allowing farmers to propagate the crops. Third-world farmers, who work hard to collect seed in order to grow their next year's crop, thereby saving money, would especially be affected. Following a great outcry—in this country but especially in Europe—Monsanto decided this November to drop plans to market "terminator" seeds. This decision should help the image of Monsanto, which had been bearing the brunt of the consumer backlash against GM crops in Europe.**

A more advanced technology is being developed [*Nature Biotechnol* 17: 1054, 1999] that will allow farmers to activate an encoded resistance in a plant by applying a chemical to the field (in effect, "turning on" the plant's genetic protection—only if pests attack).

COMMENT **No one really understands yet the function of the gene product of *BRCA1*, first discovered in 1994 and found to be responsible for increased susceptibility to familial breast and ovarian cancer. Xu and coworkers [*Nature Genet* 22: 37, 1999] have now inactivated the mouse *Brca1* gene exclusively in the cells where breast cancer normally originates—the epithelial cells lining the milk ducts. Activation of the *Brca1* gene in these cells causes genetic instability, triggering further alterations (including inactivation of the gatekeeper gene *Trp53*) that lead down the path toward tumor formation. A global knockout of the mouse *Brca1* gene [*Mol Cell Biol* 19: 7061, 1999] leads to growth retardation, infertile males, and underdevelopment of the mammary gland.**

Q **Bisphenol A is an estrogenic endocrine disruptor, present in the lining of some cans of foods. On TV I**

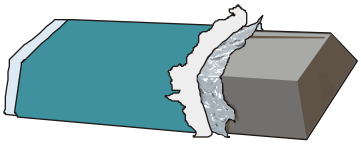
heard that bisphenol-A brings on early puberty? Should we be concerned?

A *This was a report on bisphenol A-treated mice* [***Nature*** **401:** 763, 1999]*, not a human study. Pregnant CF-1 mice on gestational days 11 to 17 were given oral doses of 2.4 µg of bisphenol A per kg body weight. This endocrine disruptor was shown to alter postnatal growth and reproductive function in females that had been exposed in utero. Although this dose was described by the authors as "equivalent to that found in the environment," I am certain that many would challenge this statement. In issue #3 [autumn of 1994], and many issues since, we have presented the pros and cons about whether endocrine disruptors are a real environmental threat.*

Q Belgium had an outbreak of dioxin contamination in their food. Can they expect to see serious illness or birth defects?

A *In Belgium, chickens contaminated in February 1999 with polychlorinated biphenyls (PCBs) and dioxins have exhibited PCB levels up to 250 times the tolerance level (0.2 µg/g fat). Some pigs were also contaminated, with levels up to 75 times the tolerance level, and no cows were found to be significantly contaminated. These chickens, their eggs, and the pigs were destroyed. It has been calculated that the consumption of at least 30 to 40 meals of highly contaminated chicken meet or eggs—would be required to double one's total body burden of PCBs and dioxin* [**Nature 401:** 231, 1999]. *And, even in this worst-case scenario, this doubling would still be more than 100 times lower than the Yusho (Japan) and Seveso (Italy) accidents of PCB and dioxin contamination, respectively. This doubling of body burden would be in the range that had been measured in subjects eating a lot of contaminated seafood in the Great Lakes area in the 1980's.*

Good News for Chocoholics



Work presented at the 1999 American Chemical Society's annual meeting revealed that chocolate is full of antioxidants [the concept of antioxidants was the leading article in issue #5, Spring 1995, of our NewsLetter]. A single chocolate candy bar (40 grams) contains more than 300 mg of polyphenols, equivalent to a day's worth of fruits and vegetables eaten in a normal diet. A single candy bar of dark chocolate contains two days' worth! Flavonoids, also present in chocolate, were shown in a test tube to help neutralize low density lipoprotein (the "bad cholesterol"). Of course, the bad news is that chocolate is high in fats and calories, and this fact must be titrated against the good news.

The New Field of 'Comparative Genomics'

Sometime around 165 million years ago, a modest rat-sized ancestral mammal, probably somewhere in Eurasia, evolved from lizard or turtle ancestors. With squared forelimbs, this animal could travel on land further and more quickly. These mammals remained diminutive, at the feet of dinosaurs, for another 100 million years. Then, a sudden extinction of the dinosaurs 63 to 66 million years ago (probably due to a meteorite causing a "nuclear winter") led to a worldwide ecological vacuum that became backfilled by the mammalian radiations.

All mammals contain somewhere between 60,000 and 150,000 genes, arranged in a linear order along their chromosomes—having a total length of probably 2.8 billion to 3.6 billion base pairs of DNA. Among mammals, the lowest known chromosome number is the Indian muntjac (3 pairs) and the highest known number is the black rhinoceros (67 pairs). Gene maps have been constructed in the human, mouse, rat, and about 30 other mammalian species. We now know that gene order is maintained—to some degree—in all species examined. So, if the order of genes is *A-B-C-D-E-F-G-H-J-K* in one species, the order may be *A-B-(F-E-D-C)-G-H-J-K* in a second species (in which an inversion has occurred), and the order on three chromosomes in a third species may be *A-B-C-D*, *G-F-E*, *H-J-K* (in which parts of the original chromosome have dispersed to locations on three different chromosomes).

Thus, once the human genome sequencing is completed, this will make the mouse genome sequencing easier, the rat genome sequencing even easier yet, and so forth with the pig, horse, cow and sheep. This field is called "comparative genomics" [*Science* **286:** 458, 1999]. Knowledge of the genome templates of each of many species will help us in two ways. **[a]** The information will resolve and interpret patterns of evolving genome organization from ancestral species. **[b]** These data will be a useful resource for locating the genetic determinants of heritable characteristics, behaviors and phenotypes**;** diseases—as diverse as osteoporosis in the horse, to infertility in the okapi, to increased susceptibility to viral infections in the mink—should become better understood more quickly, and drugs might be designed to aid such animals, thereby indirectly helping humans and worldwide economy in the long run.

Only in America.....do drugstores make the sick walk all the way to the back of the store to get their prescriptions while healthy people can buy cigarettes at the front.

Interface is supported by NIH grant # ES06096 from the National Institute of Environmental Health Sciences, and is published by the Center for Environmental Genetics, **Marshall Anderson, PhD, Director**

INTERFACE:
Genes and the Environment
Center for Environmental Genetics
University of Cincinnati
Cincinnati, Ohio 45267-0056
Fax: 513-558-0925 and 513-821-4664
E-mail: dan.nebert@uc.edu

Non-Profit
U.S.Postage
PAID
Permit No. 133
Cincinnati OH

How I Learned to Stop Worrying and Love DNA

Hail to thee O molecule
and mighty nucleotide
Wherein our fleshly destinies
Indelibly reside.

Omnipotent, ubiquitous
In phylogenetic span
Each pleiomorphic fantasy
Evolves within thy plan.

Thy codons are as poetry
As writ by greatest Pen
In base sequential mysteries
Which chromatins defend.

Two sugar phosphate helices
Pyrimidines between
With purines neatly organized
Enumerate each gene.

Thy messages are manifold
Their syntax proteinaceous
Their grammar stereotaxic and
Their grip on life tenacious.

Unwind, uptake and replicate
Until mitotic knell
Transcribe, translate, communicate!
So cell may nurture cell.

And so, with simple gratitude
I wish to close this poem;
Be it ever so humble
There's no place like genome.

Barbara S. Giesser. NEJM 295: 345, 1976

CEG - SPONSORED SPEAKERS

Craig L. Hanis, PhD
Professor, Human Genetics Center
University of Texas Health Sciences Center at Houston
November 3 1999 ***"Genetics of diabetes among Mexican Americans: Identification of a gene"***